Aquatic Toxicology 98 (2010) 230–244

Contents lists available at ScienceDirect

Aquatic Toxicology

journal homepage: www.elsevier.com/locate/aquatox

A transcriptomics-based biological framework for studying mechanisms of endocrine disruption in small fish species

Rong-Lin Wang[a,*], David Bencic[a], Daniel L. Villeneuve[b], Gerald T. Ankley[b], Jim Lazorchak[a], Stephen Edwards[c]

[a] USEPA, Ecological Exposure Research Division, National Exposure Research Laboratory, 26 W Martin Luther King Dr., Cincinnati, OH 45268, USA
[b] USEPA, Mid-Continent Ecology Division, National Health and Environmental Effects Research Laboratory, 6201 Congdon Boulevard, Duluth, MN 55804, USA
[c] USEPA, National Health and Environmental Effects Research Laboratory, 109 TW Alexander Drive, Research Triangle Park, NC 27711, USA

ARTICLE INFO

Article history:
Received 25 November 2009
Received in revised form 9 February 2010
Accepted 16 February 2010

Keywords:
Fish
Endocrine
Transcriptional
Signaling
Network
Pathway

ABSTRACT

This study sought to construct a transcriptomics-based framework of signal transduction pathways, transcriptional regulatory networks, and the hypothalamic-pituitary gonadal (HPG) axis in zebrafish (*Danio rerio*) to facilitate formulation of specific, testable hypotheses regarding the mechanisms of endocrine disruption in fish. For the analyses involved, we used data from a total of more than 300 microarrays representing 58 conditions, which encompassed 4 tissue types from zebrafish of both genders exposed for 1 of 3 durations to 10 different test chemicals (17α-ethynyl estradiol, fadrozole, 17β-trenbolone, fipronil, prochloraz, flutamide, muscimol, ketoconazole, trilostane, and vinclozolin). Differentially expressed genes were identified by one class *t*-tests for each condition, and those with false discovery rates of less than 40% and treatment/control ratios ≥1.3-fold were mapped to orthologous human, mouse, and rat pathways by Ingenuity Pathway Analysis to look for overrepresentation of known biological pathways. To complement the analysis of known biological pathways, the genes regulated by approximately 1800 transcription factors were inferred using the ARACNE mutual information-based algorithm. The resulting gene sets for all transcriptional factors, along with a group of compiled HPG-axis genes and approximately 130 publicly available biological pathways, were analyzed for their responses to the 58 treatment conditions by Gene Set Enrichment Analysis (GSEA) and its variant, Extended-GSEA. The biological pathways and transcription factors associated with multiple distinct treatments showed substantial interactions among the HPG-axis, TGF-beta, p53, and several of their cross-talking partners. These candidate networks/pathways have a variety of profound impacts on such cellular functions as stress response, cell cycle, and apoptosis.

Published by Elsevier B.V.

1. Introduction

Endocrine-disrupting chemicals (EDCs) are broadly defined as "exogenous agents that interfere with the synthesis, secretion, transport, binding, action, or elimination of natural hormones in the body responsible for the maintenance of homeostasis, reproduction, development, and/or behavior" (Kavlock et al., 1996). A more pragmatic, albeit narrower, definition of EDCs is compounds which interfere with the vertebrate hypothalamic-pituitary–gonadal (HPG) axis, leading to abnormal function and development (US EPA, 1998). Common sources of environmental EDCs include effluent from sewage treatment plants, agricultural runoff, and industrial discharges. Since the mid-1990s, there have been increasing concerns regarding EDCs and their potential harmful effects on humans and wildlife, including fish (WHO, 2002). Results so far from human epidemiological studies are largely inconclusive (WHO, 2002; Janssen et al., 2007; Hauser et al., 2007), although laboratory animal studies have demonstrated a linkage between low-dose EDC exposures and effects on a variety of reproductive and developmental endpoints (Melnick et al., 2002; WHO, 2002). In the field, relatively clear instances of impact have been found in individual fish exposed to estrogenic EDCs, or even an entire wild fish population subjected to a low level, chronic exposure to the potent estrogen 17α-ethynyl estradiol (Kidd et al., 2007). Further understanding of population-level impacts of EDCs in fish is, however, dependent on an enhanced knowledge of EDC modes

Abbreviations: DEG, differentially expressed gene; EDC, endocrine-disrupting chemical; EE2, 17α-ethynyl estradiol; E-GSEA, Extended Gene Set Enrichment Analysis; FAD, fadrozole; FIP, fipronil; FLU, flutamide; GSEA, Gene Set Enrichment Analysis; HMR, human–mouse–rat; HPG, hypothalamic-pituitary–gonadal; IPA, Ingenuity Pathway Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; MOAs, modes of action; KET, ketoconazole; MUS, muscimol; PRO, prochloraz; TF, transcription factor; TRB, 17β-trenbolone; TRI, trilostane; VIN, vinclozolin.

* Corresponding author. Tel.: +1 513 569 7862; fax: +1 513 569 7115.
E-mail address: wang.rong-lin@epa.gov (R.-L. Wang).

0166-445X/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.aquatox.2010.02.021

of action (MOAs), and development of mechanism-based indicators suitable for field work that enable linkage of exposure to adverse effects at both individual and population levels.

Over the years, there has been significant progress in elucidating the MOAs for a range of HPG-active EDCs. These often involve chemical interactions with nuclear receptors and subsequent activation or suppression of their target genes (Choi and Lee, 2004; Manning, 2005; Tabb and Blumberg, 2006; Ankley et al., 2009). Perhaps the best-known examples in this regard are androgen and estrogen receptors in liver, gonad, and other responsive tissues. Other EDC targets that have been comparatively well studied are steroidogenic enzymes, including some in the cytochrome P450 family (e.g., CYPs 11, 17, 19 or 21) and hydroxysteroid dehydrogenases (HSDs 3, 11,17, and 20) in gonadal tissue (Sanderson, 2006). Other possible endocrine MOAs that have received less attention include alterations of receptor activities without ligand binding, modulation of neuroendocrine regulation via interactions with neurotransmitter receptors (e.g., GABA or dopamine receptors), modulation of co-activators to receptors, changes in nuclear receptor turnover, and DNA methylation (Tabb and Blumberg, 2006). Little is known at present about the wider biological context under which these MOAs operate, including their regulatory control by signal transduction pathways and transcription factor (TF) networks leading to downstream manifestations of phenotypic changes at various levels of biological hierarchy. A TF network is defined here as a group of genes with ordered interactions and regulated by a common TF.

Elucidation of signaling pathways and TF networks affected by EDCs could be successfully addressed by examining transcriptomic responses in species such as zebrafish (*Danio rerio*). Among the several fish species with a sequenced genome, zebrafish is arguably one of the best model systems for ecotoxicology. It has relatively abundant genomic resources such as genetic maps, mutants, and markers available (Sprague et al., 2006), and a fairly recent divergence from fathead minnow (*Pimephales promelas*), one of the best-studied species in aquatic toxicology (Ankley and Villeneuve, 2006). Yet despite years of effort, less than half of identified protein-coding genes in zebrafish are annotated, with only 100 plus curated biological pathways, mostly metabolic, openly available (Kyoto Encyclopedia of Genes and Genomes; KEGG, www.genome.jp/kegg). The relatively sparse genome annotations not only make it difficult to frame results of toxicogenomics studies in a meaningful biological context, but also impede the formulation of new hypotheses.

Construction of a biological framework, defined as a set of signaling pathways and TF networks prioritized by their statistical associations with chemical treatments, would expand the scope of investigations of EDC regulatory mechanisms from a relatively narrow view of known HPG-axis-related targets to multiple well-annotated pathways and a still greater number of tentative (experimentally unconfirmed) networks, enabling new hypotheses to be formulated and tested for the mechanisms underlying specific toxic effects. From a systems biology perspective, signaling pathways and TF networks are at the center of a complex biological system. A small number of generally well-conserved signaling pathways across animal species ultimately rely on TFs varying in activity and specificity for sufficient regulatory complexity, which is likely responsible for much of organismic complexity in phenotypes and adaptations throughout animal kingdom (Levine and Tjian, 2003; Pires-daSilva and Sommer, 2003; Chen and Rajewsky, 2007; Vaquerizas et al., 2009). As such, signal transducers and TFs provide critical links between chemical exposures and resultant toxic effects manifested at various levels of biological hierarchy, from molecular to organismic.

A mechanistic understanding of exposure and effect at the level of signal transduction and transcriptional regulation is significant for several reasons (Ankley et al., 2009). As upstream triggers of molecular regulatory cascades, signal transducers and TFs may serve as more informative and reliable molecular indicators for exposure assessment than target genes downstream because of their mechanistic control of phenotypes and generally less variable expression (Lim et al., 2009; Vaquerizas et al., 2009). Metabolism-based computational modeling could also be made more realistic by incorporating the aspect of gene regulatory control of the enzymes involved, facilitating a shift in toxicity testing from *in vivo* apical responses to short term *in vitro* testing and predictive toxicology (NRC, 2007). Mechanistically based molecular indicators would also allow for improved extrapolation of effects across species, biological levels of organization, and diverse chemical structures. Finally, given the pleiotropic nature of signal transducers and TFs, organismic end points explicitly mapped to specific toxicity mechanisms may be developed by generating gene knockout mutants in targeted pathways. An ensuing greater efficiency and accuracy in the assessment of both EDC exposure and hazard would improve the overall risk assessment process (Ankley et al., 2010; ECETOC, 2007).

Reverse engineering of gene networks refers to a process through which a network of putative gene interactions is inferred based on the observed gene expression patterns across a range of experimental perturbations (Bansal et al., 2007). Conceptually, genes belonging to a common biological pathway should respond transcriptionally to experimental perturbations in a coordinated fashion. Using annotated TFs as hub genes (i.e., TF regulators of interest for anchoring networks by a reverse-engineering algorithm), TF regulatory networks could then be inferred based on changing gene expression profiles using computational algorithms. Gene network construction has been under intensive study in the past decade, with significant advancements made in algorithms ranging from simple coexpression clustering to more sophisticated neural networks, Bayesian networks, relevance networks, and graphic Gaussian models (Werhli et al., 2006; Blais and Dynlacht, 2005). The performance of these algorithms depends on a combination of factors, including the scale and complexity of networks, and quality of input data. Small scale and simple networks can be reverse-engineered more precisely than genome-scale large networks, where resolving the direct regulatory interactions among all the gene nodes remains a challenge (Hache et al., 2009; Baralla et al., 2009; Scheinine et al., 2009; Kim et al., 2009). For our purpose of constructing a small fish biological framework of signaling pathways and TF networks, however, it is important to note that gene membership of a network and its statistical association, or lack thereof, with EDC treatments are more important than unequivocally defining gene connectivities (Edwards and Preston, 2008).

This study is part of a larger integrated project investigating the mechanisms of endocrine disruption by several chemicals with known or hypothesized impacts on HPG-axis function (Ankley et al., 2009). One of the goals of the project is to discover mechanistically based biomarkers with utility for risk assessment and develop modeling approaches for predicting adverse outcomes. As part of this goal, the present study seeks to construct a transcriptomics-based biological framework composed of statistically prioritized signaling pathways and TF networks in order to formulate specific hypotheses for investigating EDC MOAs. The current study has five objectives: (1) identify from better annotated human–mouse–rat (HMR) genomes a set of biological pathways relevant to EDCs; (2) reverse-engineer genome-wide zebrafish TF networks; (3) use clustering analysis to characterize the interrelationships among TF networks, KEGG pathways, and a gene collection associated with the HPG-axis (hereafter referred to as "HPG-axis compiled" where appropriate); (4) prioritize TF networks, KEGG and HPG-axis compiled by statistically associating them with EDC treatments; and (5) demonstrate potential applications of candidate pathways/networks in formulating new hypotheses for studying

EDC mechanisms. Briefly, differentially expressed genes (DEGs) were identified for individual microarray-based treatments and mapped to orthologous HMR pathways using Ingenuity Pathway Analysis (IPA, www.ingenuity.com). Additionally, hundreds of TF networks were reverse-engineered computationally using an algorithm based on relevance networks. The interrelationships among TF networks, KEGG pathways, and HPG-axis compiled were evaluated by clustering analysis. Gene Set Enrichment Analysis (GSEA, Subramanian et al., 2005) and its variant, Extended-GSEA (E-GSEA, Lim et al., 2009) were then used to determine statistical associations between these gene groups organized as individual pathways/networks and the various treatment conditions. Finally, a small set of well-defined signaling pathways and their peripheral TF networks are described as possible candidates for further investigations into EDC MOAs.

2. Materials and methods

Zebrafish exposures were conducted using 10 chemicals with differing known/hypothesized MOAs within the HPG-axis: 17α-ethynyl estradiol (EE2), fadrozole (FAD), 17β-trenbolone (TRB), fipronil (FIP), prochloraz (PRO), flutamide (FLU), muscimol (MUS), ketoconazole (KET), trilostane (TRI), and vinclozolin (VIN) (Ankley et al., 2009). Details and rationale regarding experimental design, fish exposure, and gene expression profiling including microarray data preprocessing are presented elsewhere (Wang et al., 2008a,b; Ankley et al., 2009; Villeneuve et al., 2009). Briefly, reproductively mature male and female zebrafish were exposed to a continuous flow of test chemical (two different, analytically confirmed concentrations and a control), delivered in water (with no solvent), for 24, 48, or 96 h. At the end of each exposure period, fish were sacrificed in a buffered solution of tricaine methanesulfonate (MS-222; Finquel, Argent, Redmond WA, USA) and tissues including gonads, liver, and brains (with the pituitary gland and hypothalamus) were collected. Total RNA isolated from selected tissue samples was labeled and hybridized to microarrays by an Agilent certified contract laboratory (Cogenics, Morrisville, North Carolina 27560, USA). Expression profiling in zebrafish was achieved using Agilent two-color zebrafish microarrays (G2518A and G2519F, Agilent Technologies, Santa Clara, CA 95051, United States). Data from 300 plus microarrays, representing 58 conditions (Table 1) encompassing the 10 chemicals, 4 tissue types, and 3 exposure durations, in both male and female zebrafish were analyzed as part of the present investigation.

This investigation employed two complementary approaches for discovery of EDC-responsive biological pathways: identification of DEGs for individual treatments followed by mapping to orthologous HMR pathways using IPA software; and linking HPG-axis compiled, KEGG pathways, and reverse-engineered TF networks, all zebrafish specific, to these treatments by GSEA. The associations between TF networks and the treatments were also determined by E-GSEA (described in detail below). HMR pathways were not subjected to GSEA because the proprietary nature of IPA makes their statistical analysis outside the IPA framework very difficult. Data analysis procedures are summarized in Fig. 1.

2.1. Mapping DEGs to HMR pathways

Text output files from Agilent Feature Extraction software were imported into GeneSpring GX 10 (Agilent Technologies). Due to the use of two different Agilent microarray formats over the course of the study, two separate projects had to be constructed within GeneSpring, the first containing data from the EE2, FAD, and TRB experiments, conducted using G2518A (1 × 22K microarray format), and the second with data from all other chemicals, with G2519F (4 × 22K microarray format). A total of 16 dye swap controls were excluded from this analysis. Since LOWESS normalization is carried out for individual microarray data set by Agilent Feature Extraction software, GeneSpring data preprocessing and normalization were limited to value adjustment to a threshold, ratio computing, and logarithmic transformations, with no statistical estimation of missing data (i.e., imputation). The DEGs for individual treatments were identified by GeneSpring one class *t*-tests followed by Benjamini–Hochberg multiple test corrections (Benjamini and Hochberg, 1995). Due to the impact on false discovery rate (FDR) by small sample size of these treatments (Pawitan et al., 2005), the DEGs were selected at relaxed FDRs of less than 40%, but strengthened with an additional requirement of ≥1.3-fold change between a treatment and its control. These DEGs by treatment were exclusively used for their IPA mapping to HMR pathways only, at a *P*-value cutoff of 0.05.

Table 1
A list of 58 treatments denoted by chemical, duration of exposure, gender, and tissue type. EE2, ethynyl estradiol; FAD, fadrozole; TRB, trenbolone; FIP, fipronil; PRO, prochloraz; FLU, flutamide; MUS, muscimol; KET, ketoconazole; TRI, trilostane; and VIN, vinclozolin.

Numeric ID	Treatments	Numeric ID	Treatments
1	EE2 48 h male brain	30	KET 48 h female ovary
2	EE2 48 h male testis	31	KET 96 h female ovary
3	EE2 96 h female ovary	32	MUS 48 h male brain
4	EE2 96 h male brain	33	MUS 96 h male brain
5	EE2 96 h male testis	34	MUS 48 h female brain
6	EE2 48 h male liver	35	MUS 96 h female brain
7	FAD 48 h female ovary	36	MUS 96 h female ovary
8	FAD 48 h female brain	37	PRO 48 h female brain
9	FAD 96 h female brain	38	PRO 96 h female brain
10	FAD 96 h female ovary	39	PRO 48 h male testis
11	FAD 96 h male brain	40	PRO 96 h male testis
12	FAD 24 h female ovary	41	PRO 48 h female ovary
13	FIP 48 h male brain	42	PRO 96 h female ovary
14	FIP 48 h female brain	43	TRB 96 h female brain
15	FIP 48 h male testis	44	TRB 96 h female ovary
16	FIP 96 h male testis	45	TRB 24 h female ovary
17	FIP 48 h female ovary	46	TRB 48 h female ovary
18	FIP 96 h female ovary	47	TRB 96 h female liver
19	FLU 48 h male testis	48	TRI 24 h male testis
20	FLU 96 h male testis	49	TRI 48 h male testis
21	FLU 48 h female ovary	50	TRI 96 h female ovary
22	FLU 96 h female ovary	51	TRI 96 h female ovary low
23	FLU 24 h male testis	52	TRI 96 h male testes low
24	KET 96 h male brain	53	TRI 96 h male testes
25	KET 96 h female brain	54	VIN 48 h female ovary
26	KET 96 h male liver	55	VIN 24 h male testis
27	KET 96 h female liver	56	VIN 48 h male testis
28	KET 96 h male testis	57	VIN 96 h male testis
29	KET 24 h female ovary	58	VIN 96 h female ovary

2.2. KEGG and HPG-axis compiled

A group of 209 genes known to be involved in HPG-axis function was manually compiled based on published literature (Villeneuve et al., 2007). A total of 137 zebrafish pathways were downloaded from the KEGG database in October, 2008 and used in this study. As of July 2009, however, the number of zebrafish KEGG pathways totaled 129, resulting from 6 new additions and 14 deletions. The removed pathways were DRE00031, DRE00220, DRE00281, DRE00362, DRE00521, DRE00602, DRE00626, DRE00710, DRE00920, DRE00940, DRE00960, DRE01030, DRE01031, and DRE01032.

2.3. Reverse-engineering TF networks

TF networks were constructed using the ARACNE algorithm (Basso et al., 2005) which derives a transcriptional regulatory net-

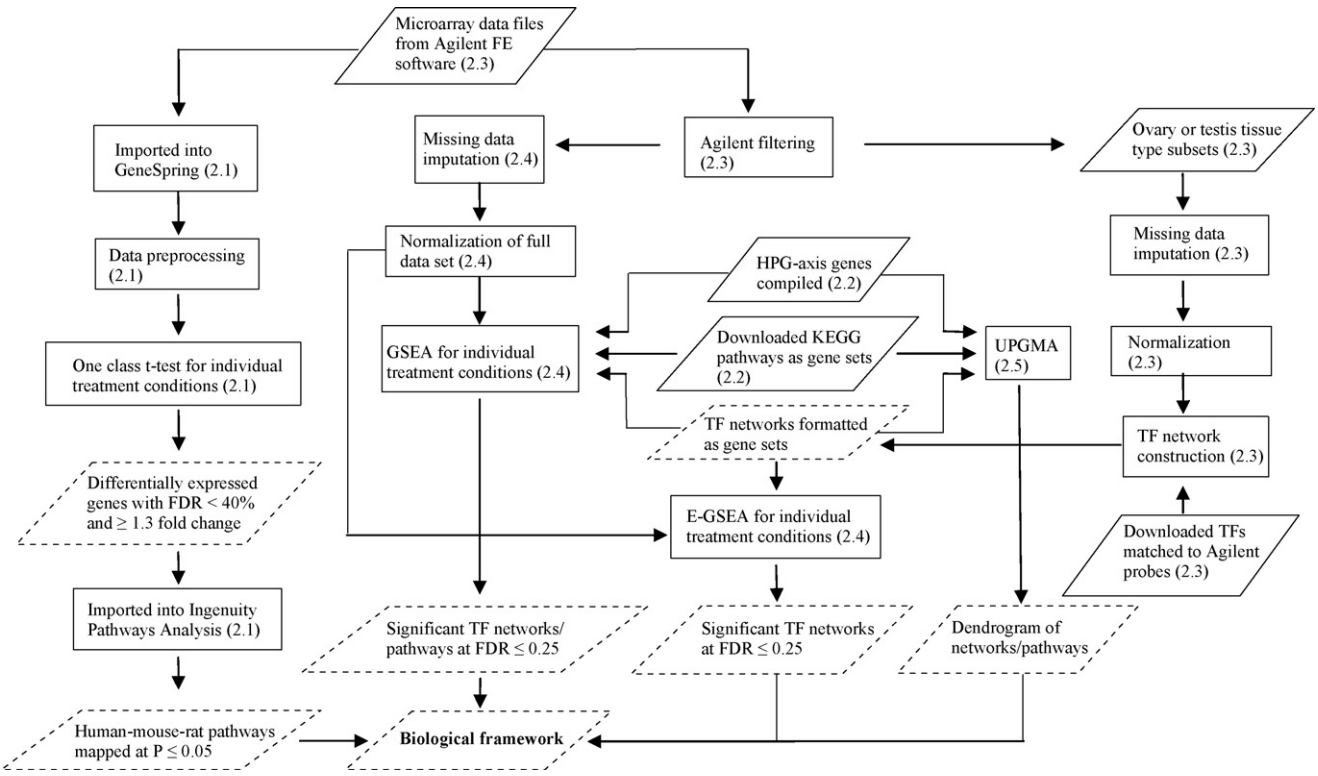


Fig. 1. The data analysis flowchart. Parallelograms are data sets used/produced in the study, with those in solid lines representing starting data for the analyses whereas those in dashed lines representing data generated during the analyses. Squares represent various methods/procedures employed. The numbers in parentheses in boxes/parallelograms refer to matched descriptions in Section 2.

work from a matrix of mutual information scores, a measure of general (including non-linear) relatedness among gene pairs. The first step was to identify annotated TFs in the Agilent zebrafish microarray as hub genes. A genome-wide collection of 1939 TFs was downloaded from www.transcriptionfactor.org (June, 2008). They contained Ensembl (http://www.ensembl.org) peptide identifications only, and were predicted based on zebrafish genome assembly version Zv7, Release 49.7 using Pfam and Superfamily (Wilson et al., 2008). The 21,322 zebrafish protein-coding, unspliced genes from the same assembly with Ensembl Gene ID and peptide ID were downloaded from Ensembl/Biomart. The 21,495 Agilent zebrafish microarray probe sequences were mapped to the 21,322 genes in the genome by BLASTN, yielding 11,995 unique probe-gene matches at *E* values less than or equal to 1E−12. This led to a final linkage of 951 Agilent zebrafish microarray probes to TFs.

The second step of TF network construction was a preparation of microarray gene expression data sets using KNNimputer (Troyanskaya et al., 2001) and several Perl scripts we developed (www.perl.org). Since the ARACNE algorithm requires a minimum of 100 samples from the same tissue type under a variety of perturbations (Margolin et al., 2006) and, in the present study, Cyanine 5 (Cy5) and Cyanine 3 (Cy3) labeled RNA represent unique biological samples, 3 separate gene expression data sets were qualified: female ovary data in log ratios (105 samples, hereafter OvaryRatio), female ovary data in single channel intensities (205 samples excluding 5 duplicated Cy3 samples, hereafter OvaryCy5Cy3), and male testis data in single channel intensities (163 samples excluding 5 duplicated Cy3 samples, hereafter TestisCy5Cy3). After individual samples were preprocessed based on Agilent recommended filters, rProcessedSignal and gProcessedSignal from each sample were assembled according to the sample tissue type into either an ovary or a testis data set. Genes with less than 75% valid entries across all concatenated samples within a tissue type were not considered further. Missing values for the remaining genes were imputed separately for the ovary and testis data set by KNNimputer, based on the corresponding expression values of the most similar samples. Two input files were then generated from the ovary data: OvaryRatio by value adjustment to a threshold, ratio computing, and $\log_2$ transformation; and OvaryCy5Cy3 by value adjustment to a threshold, $\log_2$ transformation, per chip normalization to 75th percentile, and per gene normalization to the median of all samples. The TestisCy5Cy3 data set was generated from the testis data using the same normalization scheme as OvaryCy5Cy3. After data preprocessing, each data set had only a fraction of the original 21,495 probes/951 TF probes left: OvaryRatio and OvaryCy5Cy3, 12,699 probes containing 554 TF probes; and TestisCy5Cy3, 14,148 probes containing 690 TF probes. Since a single TF could be represented by multiple unique probes on the Agilent Microarray, the actual number of TF genes is less than the stated and variable depending upon evolving gene annotations.

The latest version of the algorithm, ARACNE2, was provided to us in October, 2008 by Manjunath Kustagi from the Andrea Califano lab at the Center for Computational Biology and Bioinformatics, Columbia University (New York City, NY 10027, USA). ARACNE2 was run on all three data sets along with their respective lists of TF probes using two Linux clusters and a Condor computer grid (http://www.cs.wisc.edu/condor). Each data set was bootstrapped 100 times and run independently, and each TF probe was used as a hub TF once within a data set. ARACNE2 configuration parameters were specifically calculated for each of the three data sets courtesy of Manjunath Kustagi. Sample arguments for running ARACNE2 were "−i gene expression data set −o input_$(Process).adj -p 5e-8 -e 0.1 -r $(Process) -s a list of TFs -l a list of TFs". It took ARACNE2 approximately 261, 178, and 441 h, respectively, to process the OvaryCy5Cy3, OvaryRatio, and TestisCy5Cy3 data sets, yielding a total of 554, 550, and 690 TF networks. The resulting 100 adjacency files, containing inferred gene interactions and their mutual

information scores, from each data set were combined by "getconsensusnet.pl", provided as part of ARACNE2 package, to generate a consensus adjacency file. The consensus TF-gene networks for each of the three data sets were converted into GSEA gene matrix transposed format (gmt) by replacing Agilent probes with their gene symbols as assigned by the Human Gene Organization Gene Nomenclature Committee (http://www.genenames.org), based on Agilent zebrafish annotations. Each TF network was confirmed to contain a unique set of gene symbols. While there are probably varying degrees of conservation in the compositions of these TF networks across tissue types, due to the exploratory nature of the current study, we adopted an inclusive approach where each TF network was utilized as a gene set in GSEA or E-GSEA for all 58 treatments regardless of its original tissue.

2.4. GSEA and E-GSEA

For GSEA and E-GSEA, 2 gene expression data sets containing 290 and 291 microarrays, inclusive of all 58 treatment groups, were prepared in single channel intensities in similar steps as described in Section 2.3. These contained a total of 12,802 and 12,793 probes, respectively. Both analyses were conducted based on data from individual treatments in the 290 data set. The 291 microarray data set contains an extra dye swap array from the EE2 48 h male testis condition in order to increase its sample size for E-GSEA of this particular treatment.

A GSEA R script (http://www.r-project.org), GSEA-P-R.1.0, was downloaded from The Broad Institute at http://www.broad.mit.edu/gsea/. To enable batch processing, a separate R script wrapper, Run_Broad_GSEA.r, was written with a few minor modifications to the original GSEA R script. GSEA was conducted for each of the 58 treatment conditions using the 290 microarray gene expression data set, and gene sets made up of HPG-axis compiled, KEGG pathways, and TF networks. Each gene set needed to contain at least 15 genes, but no more than 2000, in order to be included. The test FDR was set at 0.25, a typical significance cutoff for GSEA (and E-GSEA) because of the exploratory nature of this analysis and a general lack of coherence found in most gene sets.

GSEA associates an a priori defined set of genes to a treatment by determining whether the gene set as a whole is statistically enriched among the ranked DEGs between the treatment and its control. To handle gene sets containing both up- and down-regulated genes as expected in a TF network, an Extended-GSEA (E-GSEA) procedure has been proposed (Lim et al., 2009), which incorporates the expected relationship between a TF and its regulated genes derived from a TF network. To implement this method in R, we significantly modified the original GSEA algorithm based on the GSEA R script from the Broad Institute and two demonstrative pieces of Matlab code made available to us courtesy of Wei Keat Lim (Lim et al., 2009). This modified approach divides a gene set into either a TF-activated or -repressed subset as measured by TF-gene correlation and a gene list into a chemical up- or down-regulated subgroup as measured by signal-to-noise (SNR) ratios. The relationships among these four groups of genes are then individually considered in order to capture the contributions to the enrichment statistic by both TF-activated and -repressed genes. Additional details on the algorithm are given in Lim et al. (2009). Where multiple probes exist in the expression data for a given gene, as is often the case with the Agilent zebrafish microarray, the one with the largest SNR in absolute value is retained and the rest removed. Unlike the original GSEA, the sign of the enrichment statistic for a gene set no longer indicates the up- or down-regulation of that gene set by a given treatment. Although no upper ceiling was imposed, each TF network must have at least 15 gene members in order to be included for E-GSEA. The E-GSEA approach is computationally quite expensive, taking about 137 CPU hours per treatment for 1465 TF networks on a Linux cluster with 128 Intel Itanium2 1.6 GHz processors.

2.5. Similarity comparison among TF networks, HPG-axis compiled, and KEGG pathways

To compare similarity among HPG-axis compiled, KEGG pathways, and TF networks, an R script was written to calculate pairwise Jaccard distance (Jaccard, 1901) among these pathways/networks. Given two sets of genes from two pathways to be compared, Jaccard distance (D_j) is defined as:

$$D_j = 1 - \frac{\text{number of genes in the intersection}}{\text{number of genes in the union}}$$

Two distance matrices were calculated, one for HPG-axis compiled and KEGG pathways only totaling 138, and the other for all KEGG pathways, HPG-axis compiled, and TF networks combined (1932 in total). An UPGMA (Unweighted Pair Group Method with Arithmetic mean) dendrogram was generated from each resulting distance matrix using MEGA4 (Tamura et al., 2007) and visualized in Dendroscope (Huson et al., 2007).

3. Results

A TF network consists of a hub TF, as represented by a unique probe, and all of its interacting gene targets (Margolin et al., 2006). Although treated as a regulating hub gene only once in each data set, a TF, like any other gene, could be a target of regulation in multiple networks. In this paper, a TF network is referred to by its hub TF. To improve readability, a TF will be represented by its unique upper case gene symbol only. Since TF networks were constructed with three different gene expression data sets, OvaryCy5Cy3, OvaryRa-

Table 2
Summary statistics of reverse-engineered zebrafish TF networks. Three microarray data sets, OvaryCy5Cy3, OvaryRatio, and TestisCy5Cy3, were prepared for ovary and testis tissue types in either single channel intensities or Cy5/Cy3 ratios. The data sets OvaryCy5Cy3 and OvaryRatio share identical set of TF probes.

Parameters	OvaryCy5Cy3	OvaryRatio	TestisCy5Cy3
Sample size	205	105	163
No. hub TFs	554	554	690
Network size minimum	2	1	9
Network size median	250	52	220
Network size maximum	1928	714	1046
Network size average	373	105	262
Total no. genes networked	10,357	8116	11,593
Total no. input genes	12,699	12,699	14,148
Gene networking frequency maximum	116	47	104
Gene networking frequency median	14	5	13
Gene networking frequency average	20	7	16
No. TF networks with enriched Gene Ontology biological processes[a]	48	27	65

[a] Conducted by web submission to http://discover.nci.nih.gov/gominer/htgm.jsp (Zeeberg et al., 2005).

tio, and TestisCy5Cy3, the descriptor for a TF network in the tables follows the format of DRTF (*D. rerio* transcription factor) – data set – Agilent probe ID, for example DRTFovaryRatio_A_15_P110418. Only annotations for hub TFs are listed. Individual treatments are named according to test chemical, exposure time in hours, gender, and tissue type. For a chemical treatment where two test concentrations were evaluated, the name of the lower dosage treatment ends with "low". Treatments are listed individually only in the first table and represented by numeric identifiers thereafter. Due to the size of the data sets and complexity of analyses, only summary or selected results are reported. Complete outputs from various analyses are provided as supplementary materials and are available online.

3.1. A zebrafish biological framework for studying EDC mechanisms

After preprocessing (see Section 2.3), data sets OvaryCy5Cy3, OvaryRatio, and TestisCy5Cy3 contained 554, 554, and 690 TFs respectively, out of a total of 951 annotated and compiled zebrafish TFs. About 60–80% of input genes are networked by the ARACNE algorithm (Table 2). On average, TF networks built based on the OvaryCy5Cy3 data set were the largest, followed by TestisCy5Cy3 and OvaryRatio. This appears to be a function of sample size in each data set. A single gene could be a member across multiple networks (see gene networking frequency, Table 2). Some of them, such as YBX1 and SMAD2, were highly networked (data not shown).

UPGMA clustering of TF networks, KEGG pathways, and HPG-axis compiled based on Jaccard distance offers a global view of their relative similarities (Fig. 2A and B; Supplemental Figure 1). A longer branch between two pathways in the dendrogram indicates a greater distance, and all branches are relative to scale. Apparent in the dendrogram with only KEGG pathways and HPG-axis compiled are two relatively large clusters bordered by DRE00624/DRE00640 and DRE04310/DRE04370, a number of smaller clusters, and some quite distinct individual pathways (Fig. 2A). When all 1932 pathways and networks were considered together, most of the KEGG pathways clustered into a few distinct groups with a dozen of them forming outliers scattered throughout the dendrogram (Supplemental Figure 1). Those KEGG pathways known to interact with one another tend to have more genes in common and are therefore clustered closer together. Two such examples are the cluster containing N-glycan biosynthesis (DRE00510) and several sulfate metabolic pathways (Supplemental Figure 1), and the cluster containing the TGF-beta signaling pathway (DRE04350, Fig. 2A). Overall, TF networks tended to form local clusters defined by one of the three data sets.

Pathways and TF networks were evaluated for their relevance to different mechanisms of endocrine disruption by two complementary approaches: mapping DEGs from individual treatments to HMR pathways by IPA, and conducting GSEA/E-GSEA on KEGG pathways and reverse-engineered TF networks (Fig. 1). To reduce the impact of widely variable FDRs among DEG sets from different treatments (data not shown) on HMR pathway mapping, an arbitrary criterion of less than 40% FDR and ≥1.3-fold change between a treatment and its control was imposed on each set of mapped DEGs. This eliminated DEGs from 18 of 58 treatment conditions. In addition, of 137 KEGG pathways, only 33 met the minimum size standard of 15 to be included in GSEA. Among the disqualified pathways, most were too small in size and the rest were dropped due to unsuccessful mappings between KEGG pathways and Agilent probes. For the 1794 TF networks constructed, 1673 were qualified for GSEA with a network size between 15 and 2000 genes, and 1465 qualified for E-GSEA with both network size ≥15 and hub TFs present in the individual gene expression data sets after filtering. A total of 1412 pathways/networks were found to be chemical treatment-significant at a *P*-value ≤ 0.05 or FDR ≤ 0.25 (Fig. 1): 149 HMR, 30 KEGG, 1232 TF networks, and HPG-axis compiled. Among the significant TF networks were 515 across 53 treatments by E-GSEA and 1126 across 35 treatments by GSEA, with 409 networks overlapped between the two methods. Examined by the data sets, GSEA-significant networks included 541 from TestisCy5Cy3, 364 from OvaryCy5Cy3, and 221 from OvaryRatio. E-GSEA has 238 from TestisCy5Cy3, 187 from OvaryCy5Cy3, and 90 from OvaryRatio.

Table 3
Selected IPA-significant HMR pathways among a total of 149, as ranked by number of treatments, of which their DEGs are mapped to the pathways at a *P*-value cutoff of 0.05. IPA mapping of DEGs to HMR is limited to those DEGs from 40 treatments where FDRs for each set of DEG are <40% and treatment/control is ≥1.3-fold. IPA, Ingenuity Pathway Analysis; HMR, human, mouse, rat; DEG, differentially expressed gene.

Significant HMR pathways	Treatments
Clathrin-mediated endocytosis	2, 13, 19, 20, 23, 33, 36, 37, 38, 43, 46
IGF-1 signaling	19, 21, 23, 38, 43, 46, 47, 51
Huntington's disease signaling	14, 32, 33, 34, 35, 36, 46, 47
CTLA4 signaling in cytotoxic T lymphocytes	13, 14, 15, 32, 33, 36, 38, 46
TGF-beta signaling	8, 23, 43, 46, 47, 51, 56
NRF2-mediated oxidative stress response	1, 2, 6, 19, 22, 38, 46
Estrogen receptor signaling	6, 11, 15, 23, 43, 47, 51
VEGF signaling	20, 21, 22, 23, 46, 51
IL-8 signaling	22, 23, 38, 40, 47, 51
LPS-stimulated MAPK signaling	22, 23, 38, 46, 51
Jak/Stat signaling	1, 21, 23, 46, 51
PI3K/AKT signaling	20, 21, 23, 38, 46
IL-3 signaling	22, 23, 38, 46, 51
Hypoxia signaling in the cardiovascular system	14, 22, 25, 36, 51
ERK/MAPK signaling	16, 21, 38, 46, 56
IL-6 signaling	21, 22, 23, 46
IL-2 signaling	21, 23, 46, 51
P53 signaling	14, 57

3.2. Application of the transcriptomics-based biological framework in formulating new hypotheses

The zebrafish biological framework constructed in this study could be examined in a variety of ways to formulate new hypotheses for EDC regulatory mechanisms, such as by *P*-value or FDR of different combinations of treatments, network similarity in a dendrogram to a given targeted or otherwise better characterized pathway, or number of treatments under which a pathway is significant. They could also be used as gene sets in new studies to be linked to additional chemical treatments by GSEA or E-GSEA. Below we demonstrate the utility of this framework by ranking pathways/networks according to the number of chemical treatments they are associated with and examining some of the top candidates (Tables 3 and 4).

A number of top-ranked HMR pathways stand out: IGF-1, TGF-beta, Jak–Stat, PI3K, IL-2, -3, -6 (Table 3). IGF-1 and TGF-beta are significant under eight and seven chemical treatments respectively, five of which are shared between the two pathways. The p53 pathway was ranked relatively low. Table 4 lists a number of TF networks ranked high in GSEA or E-GSEA, many of which, such as ATF4, CSDC2, DPF2, TP53, are considered as top candidates by both algorithms. In general, for a given significant TF network, GSEA tended to find a greater number of associated treatments than E-GSEA did. In contrast to this is the number of treatments under which p53 signaling and TGF-beta signaling (as from HMR and KEGG) are significant according to IPA and GSEA: two versus six for p53 signaling, and seven versus zero for TGF-beta signaling. This discrepancy could probably be explained by several factors such as differences in statistical algorithms between IPA and GSEA, multi-level regulation of p53 and TGF-beta, or even composition

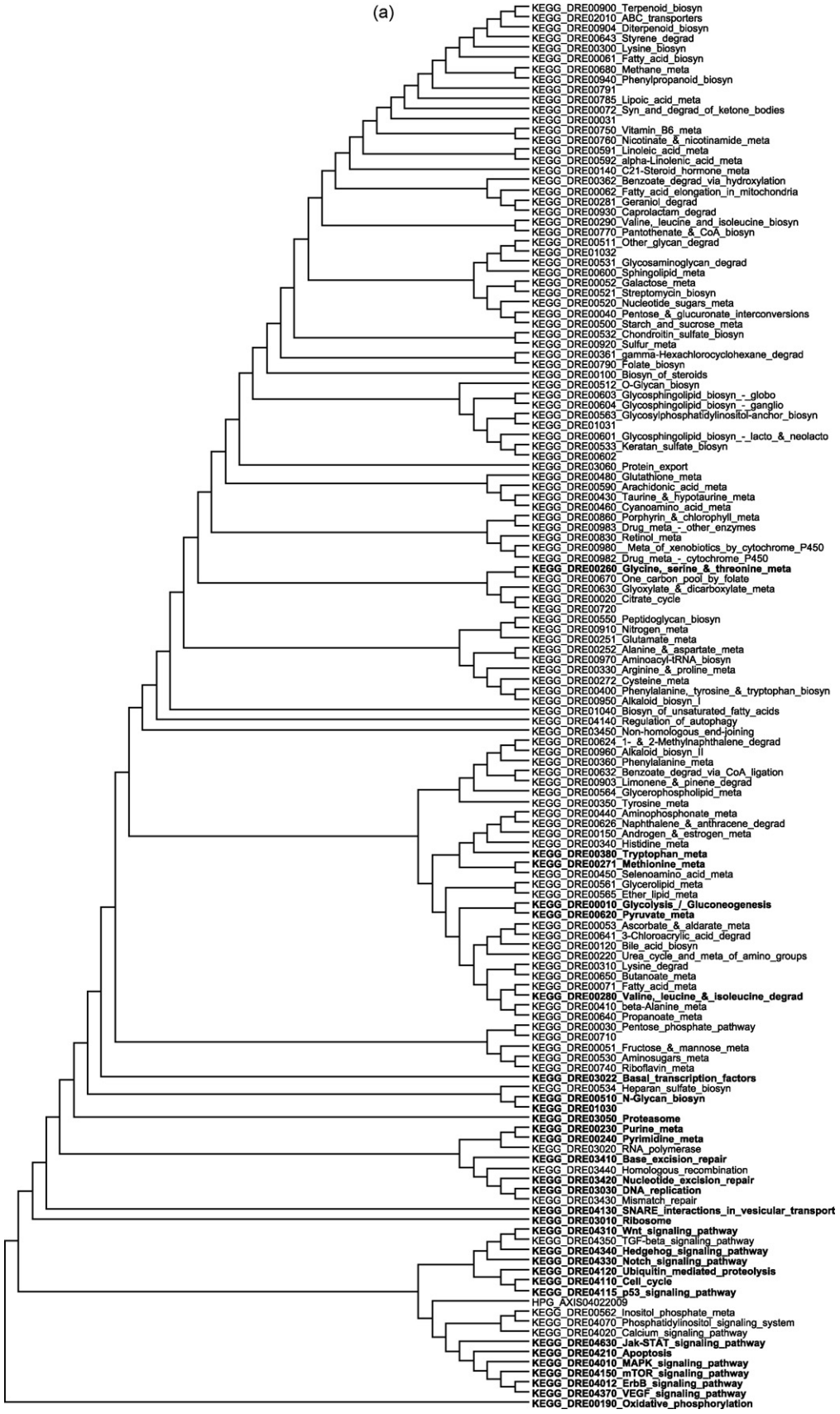
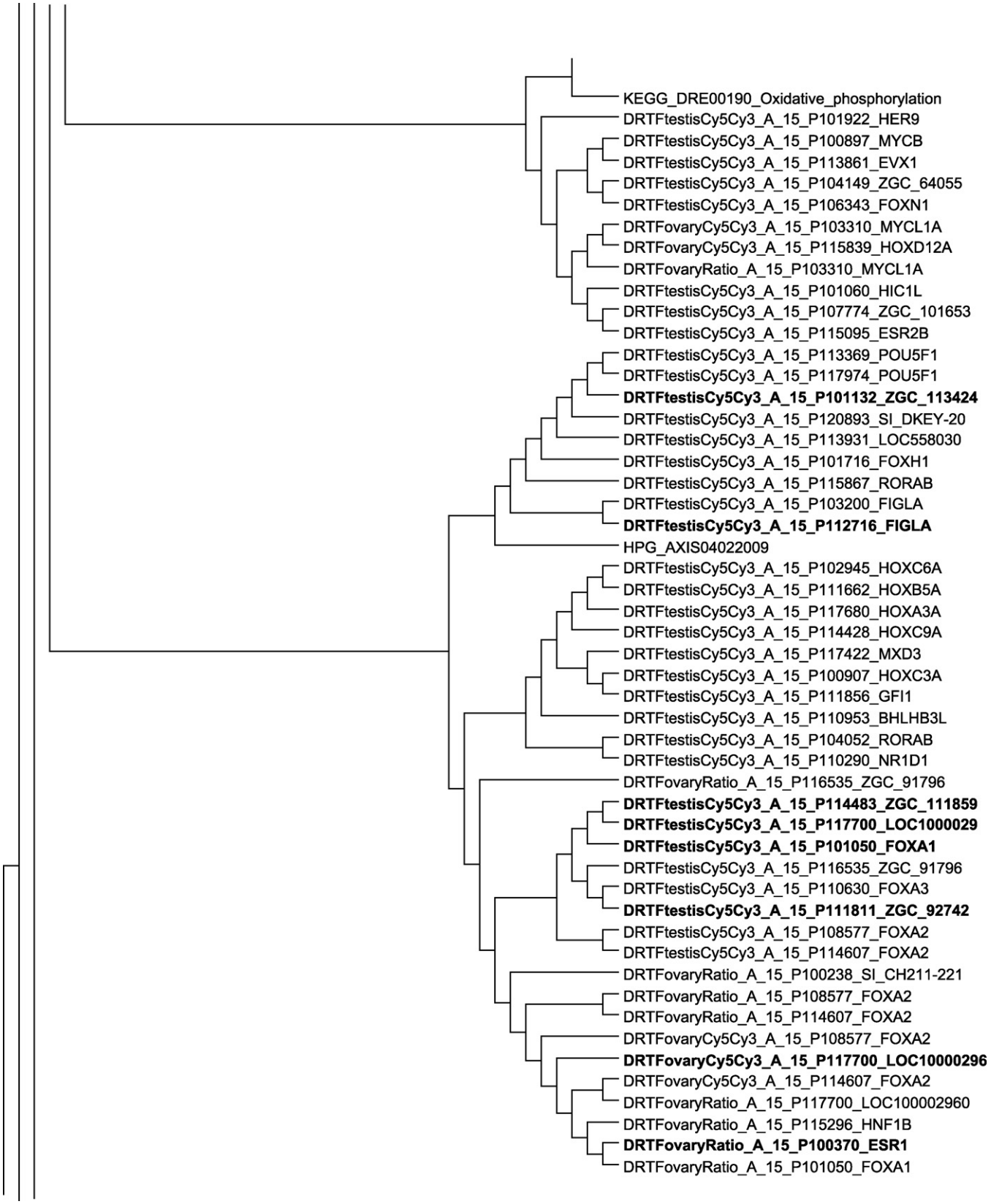


Fig. 2. (A and B) Nodes of networks/pathways adjoining HPG-axis compiled (HPG_AXIS04022009) from UPGMA (Unweighted Pair Group Method with Arithmetic mean) analysis of Jaccard distance matrix of either A: HPG-axis compiled/KEGG pathways only (all 138 shown); or B: a subsection of Supplemental Figure 1 showing only HPG-axis compiled and its most closely related nodes among the 1932 in total. EDC-significant nodes according to GSEA (A) or both GSEA and E-GSEA (B) are highlighted in bold. A branch length is relative and proportional to distance between two adjoining nodes or clusters. (A) Legacy pathways are not annotated. DRE00562, and DRE04070 were disqualified for GSEA due to small pathway size. Abbreviations: biosyn., biosynthesis; degrad., degradation; meta., metabolism; (B) Hub TF probes where annotations are



available: BHLHB3L, basic helix–loop–helix domain containing, class B, 3 like; ESR1, estrogen receptor 1; ESR2B, estrogen receptor 2b; EVX1, even-skipped homeobox 1; FIGLA, factor in the germline alpha; FOXA1, forkhead box A1; FOXA2, forkhead box A2; FOXA3, forkhead box A3; FOXH1, forkhead box H1; FOXN1, forkhead box N1; GFI1, growth factor independent 1; HER9, hairy-related 9; HNF1B, HNF1 homeobox b; HOXA3A, homeo box A3a; HOXB5A, homeo box B5a; HOXC3A, homeo box C3a; HOXC6A, homeo box C6a; HOXC9A, homeo box C9a; HOXD12A, homeo box D12a; LOC100002960, similar to LOC733301 protein; LOC100002960, similar to LOC733301 protein; LOC100002960, similar to LOC733301 protein; LOC558030, transcription factor IIIA like; MXD3, MAX dimerization protein 3; MYCB, myelocytomatosis oncogene b; MYCL1A, v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian) a; NR1D1, nuclear receptor subfamily 1, group d, member 1; POU5F1, POU domain, class 5, transcription factor 1; RORAB, RAR-related orphan receptor A, paralog b.

Table 4
Selected significant KEGG pathways and TF networks as ranked by the number of treatments under which a pathway/network is significant at a FDR ≤ 0.25. For comparison, a network ranked high in either GSEA or E-GSEA is also cross-referenced under the other method if found significant. Additional key TFs biologically relevant to those high-ranking pathway/networks are also included. There are a total of 30 KEGG and 1126 TF networks found significant by GSEA, 515 TF networks by E-GSEA. The significant TF networks between the two approaches have an overlap of 409. If the networks of a hub TF built from multiple data sets are all significant, they are listed in an abbreviated format. For example, A_15_P110418 (ovaryS/ovaryR/testisS) represents DRTFovaryCy5Cy3_ A_15_P110418, DRTFovaryRatio_ A_15_P110418, and DRTFtestisCy5Cy3_ A_15_P110418, each of which is significant under its corresponding group of treatments delineated by a forward slash.

Significant pathways and networks	Annotations	Treatments
GSEA		
A_15_P110418 (ovaryS/ovaryR/testisS)	ATF4 activating transcription factor 4	6, 13, 18, 29, 30, 33, 35, 54/13, 18, 29, 30, 33/17, 30, 33, 50
A_15_P115702 (ovaryS/ovaryR/testisS)	ATF7B, activating transcription factor 7b	17, 38, 50/38, 50/17, 38, 50
A_15_P117307 (ovaryR/testisS)	CSDC2, cold shock domain containing C2, RNA binding	2/8, 18, 29, 30, 33
A_15_P111038 (ovaryS/ovaryR/testisS)	DPF2L, D4 zinc and double PHD fingers family 2, like	17, 21, 33, 47/17, 50/2, 17, 18, 30, 33, 34, 35, 46
A_15_P104909 (ovaryS/ovaryR/testisS)	DPF2, D4 zinc and double PHD fingers family 2	17, 38, 50/17, 21, 23, 38, 50/6, 17, 38, 50
A_15_P116913 (ovaryS/ovaryR/testisS)	DPF2	38, 50/21, 23, 38, 50/17, 21, 38
A_15_P107347 (ovaryS/ovaryR/testisS)	MEF2D myocyte enhancer factor 2d	17, 21, 38, 50/21, 23, 38, 50/17, 21, 23, 38, 42, 50
A_15_P121463 (ovaryS/ovaryR/testisS)	MEF2D	38, 50/17, 23, 38, 50, 56/17, 38, 50
A_15_P105566 (ovaryS/ovaryR/testisS)	MYBL2 myeloblastosis oncogene-like 2	17, 50/17, 50/17, 21, 23, 38, 41, 42, 50, 56
A_15_P106350 (ovaryS/ovaryR/testisS)	MYBL2	6, 17, 21, 50/38, 50/17, 21, 23, 38, 50, 56
A_15_P119126 (ovaryS/testisS)	KLF2B Kruppel-like factor 2b	50/21, 23, 37, 38, 39, 42, 50, 56
A_15_P108794 (ovaryS/ovaryR/testisS)	KLF15 Kruppel-like factor 15	21, 38, 50/16, 21, 23, 38, 42/21, 38
KEGG_PATHWAY_DRE00510	N-glycan biosynthesis	3, 4, 7, 10, 15, 26, 27, 44, 46, 56
A_15_P102464 (ovaryS/testisS)	PIAS2 protein inhibitor of activated STAT, 2	6/21
KEGG_PATHWAY_DRE04115	P53 signaling	10, 14, 15, 19, 39, 46
A_15_P112675 (ovaryS/ovaryR/testisS)	TFDP1 transcription factor Dp-1	38, 50/38, 50/38, 50
A_15_P103068 (ovaryS/ovaryR)	TGIF1 TGFB-induced factor homeobox 1	38, 50/21, 23, 38, 42, 50
A_15_P102660 (ovaryS/testisS)	TP53 tumor protein p53	6, 17/2, 6, 19, 21, 23, 38, 56
A_15_P120536 (ovaryS/testisS)	TP53	6, 50/6, 19, 21, 38
A_15_P118142 (testisS)	TP73	8
A_15_P114243 (ovaryS)	XBP1, X-box binding protein 1	6, 21, 38
A_15_P111423 (ovaryS/testisS)	XBP1	6/6
A_15_P118771 (testisS)	YBX1 Y box binding protein 1	2, 6, 17, 18, 33, 34, 46
A_15_P103153 (ovaryR/testisS)	YY1 YY1 transcription factor	33, 40/8, 13, 17, 18, 29, 30, 33, 34, 35, 46
E-GSEA		
A_15_P110418 (ovaryS/testisS)	ATF4	23, 30, 33, 45/6, 33
A_15_P115702 (ovaryS/testisS)	ATF7B	38, 50/17, 38, 50
A_15_P117307 (ovaryR/testisS)	CSDC2	30/4, 15, 22, 29, 30, 33
A_15_P105678 (ovaryR)	DPF1, D4 zinc and double PHD fingers family 1	2, 15, 33, 47
A_15_P116913 (ovaryS/testisS)	DPF2	38/5, 38, 47
A_15_P104909 (ovaryS/testisS)	DPF2	38, 50/38, 50
A_15_P111038 (testisS)	DPF2L	2, 6
A_15_P121463 (ovaryR)	MEF2D	21
A_15_P102464 (ovaryS/ovaryR/testisS)	PIAS2	4, 6, 44/2/4, 6
A_15_P112675 (ovaryS/testisS)	TFDP1	4/4, 38, 50
A_15_P102660 (ovaryS)	TP53	6
A_15_P120536 (ovaryS/testisS)	TP53	39/2, 4, 27
A_15_P111423 (ovaryS/testisS)	XBP1	4, 6, 26/4, 6, 50
A_15_P114243 (ovaryS/testisS)	XBP1	6, 45, 50/4, 27, 50
A_15_P118771 (ovaryS/testisS)	YBX1	2/10
A_15_P103153 (testisS)	YY1	11

of these curated pathways themselves. IPA mapping uses Fisher's Exact Test to evaluate non-random associations between DEGs and annotated canonical HMR pathways orthologous to zebrafish genes, while GSEA/E-GSEA is a permutation-based approach to link zebrafish-specific KEGG pathways/TF networks to treatments. The regulation of p53 and TGF-beta signaling occurs at the transcriptional, translational, and post-translational levels whereas the microarray studies performed here only evaluate the first. As to a comparison between HMR pathways and TF networks, it should be noted that the genome-wide TF networks constructed in this study are tentative in nature, and subject to limitations of current reverse-engineering algorithms. In many cases, features included in our TF networks also do not have any counterparts in HMR pathways, which are available only in very limited numbers and more suitable for a general guidance of fundamental biological functions impacted by EDCs. In effect, TF networks identified in Table 4 complement the HMR pathways in Table 3 by fully leveraging the transcriptional changes and providing specific inferences as to the implications for signaling and metabolism.

The networks of individual TFs in KEGG pathway p53 signaling (KEGG DRE04115) and TGF-beta signaling (KEGG DRE04350) were further examined for their responses to the chemical treatments. For p53 signaling, 35 of 50 genes were mapped to Agilent probes with two (TP53 and TP73) curated as TFs. The networks of TP53 and TP73 were impacted by a combined total of 13 treatments (Table 4), with EE2 48 h male liver affecting the largest number of networks. Only two treatments, FLU 48 h male testis and PRO 48 h male testis, had a significant impact on both TP53/TP73 TF networks and KEGG p53 signaling pathway. For an illustration, Fig. 3 shows one of the TP53 networks and two of its closest network neighbors headed by MXI1 and NEIL3, both of which are involved in tumorigenesis or DNA repair (Rottmann and Lüscher, 2006; Takao et al., 2009) thus functionally relevant to TP53. For the TGF-beta signaling pathway, 57 of 80 members were mapped to Agilent probes, 9 of which are TFs (ID2A, ID2B, MYCB, PITX2A, SMAD1, SMAD2, SMAD3A, SMAD5, and TFDP1L). An additional four annotated zebrafish TFs involved in the pathway were also added (SMAD6, 7, 9, and TFDP1) to the list. Among the networks for these 13 TFs, 12 and 9, respectively, are significant by GSEA or E-GSEA, under a total of 15 treatments involving 8 of the 10 test chemicals (Table 5). Only two such treatments, FLU 24 h male testis and TRI 96 h female ovary low, also have an impact on HMR TGF-beta pathway. The three treatments with impact on the largest number of these TF networks were EE2 48 h male liver, TRI 96 h female ovary, and PRO 96 h female brain.

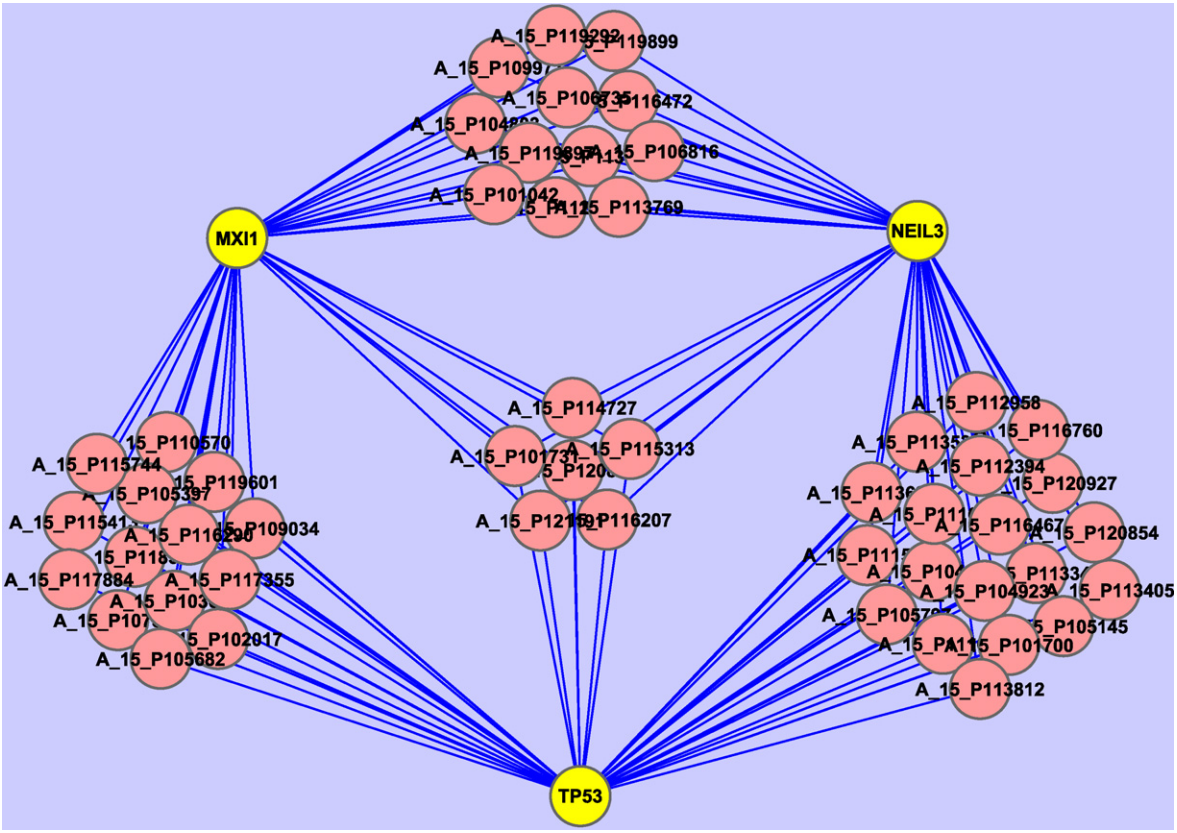


Fig. 3. The interactive TF networks of TP53 (DRTFtestisCy5Cy3_A_15_P120536) and two of its closest neighbors headed by MXI1 (DRTFtestisCy5Cy3_A_15_P103761) and NEIL3 (DRTFtestisCy5Cy3_A_15_P121124) as determined by network clustering. Shown are only target nodes having interactions with at least two out of the three hub TF probes.

Table 5
Thirty-eight TF networks containing 12 TFs from TGF-beta signaling pathway, and treatments under which they are significant by either GSEA or E-GSEA at FDRs ≤0.25. Out of 80 members of TGF-beta pathway, 9 are identified as TFs and mapped to Agilent probes. Four additional TFs (SMAD6, 7, 9, and TFDP1) are also added. If the networks of a hub TF built from multiple data sets are all significant, they are listed in an abbreviated format. For example, A_15_P102595 (ovaryS/ovaryR/testisS) represents DRTFovaryCy5Cy3_A_15_P102595, DRTFovaryRatio_ A_15_P102595, and DRTFtestisCy5Cy3_ A_15_P102595, each of which is significant under its corresponding group of treatments delineated by a forward slash.

Significant TF networks	Annotations	Treatments
GSEA		
A_15_P103330 (ovaryS/testisS)	ID2A inhibitor of DNA binding 2, dominant negative helix–loop–helix protein, a	6/2
A_15_P102595 (ovaryS/ovaryR/testisS)	ID2A	6, 38/23/17
A_15_P113186 (ovaryS)	ID2B	50
A_15_P110041 (ovaryS)	ID2B	50
A_15_P102524 (ovaryS/ovaryR)	MYCB myelocytomatosis oncogene b	6, 38, 50/38, 50
A_15_P100897 (ovaryS/testisS)	MYCB	6/6
A_15_P103771 (testisS)	SMAD1 mothers against decapentaplegic homolog 1	17, 50
A_15_P100787 (testisS)	SMAD1	17, 38, 50
A_15_P119732 (ovaryS/ovaryR/testisS)	SMAD2	6, 17, 38, 50/17, 38, 50/17, 38, 50
A_15_P101086 (ovaryS/ovaryR/testisS)	SMAD2	6/38/6
A_15_P120722 (testisS)	SMAD3B	2
A_15_P113997 (ovaryS/ovaryR/testisS)	SMAD5	6, 50/17/2
A_15_P111897 (testisS)	SMAD6	38
A_15_P121028 (ovaryS/ovaryR/testisS)	SMAD7	6/2, 3/6
A_15_P113640 (ovaryS/testisS)	SMAD9	50/2
A_15_P111335 (ovaryS/ovaryR/testisS)	TFDP1L transcription factor Dp-1, like	6/38, 50/6
A_15_P112675 (ovaryS/ovaryR/testisS)	TFDP1 transcription factor Dp-1	38, 50/38, 50/38, 50
A_15_P117603 (ovaryS/testisS)	TFDP1	33/6
E-GSEA		
A_15_P103330 (testisS)	ID2A	2
A_15_P110041 (testisS)	ID2B	25, 33
A_15_P102524 (ovaryS)	MYCB	6
A_15_P100897 (ovaryS/testisS)	MYCB	4, 6/11
A_15_P119732 (ovaryS/ovaryR)	SMAD2	4/31
A_15_P101086 (ovaryS)	SMAD2	4, 6
A_15_P113997 (ovaryS)	SMAD5	4, 6
A_15_P121028 (testisS)	SMAD7	11
A_15_P113640 (testisS)	SMAD9	2, 5
A_15_P111335 (ovaryS)	TFDP1L	6, 35
A_15_P112675 (ovaryS/testisS)	TFDP1	4/4, 38, 50
A_15_P117603 (ovaryS)	TFDP1	51

Table 6
Overlap between 209 genes from an incomplete, manually compiled HPG-axis and DEGs from 40 treatments where FDRs for each set of DEG are <40% and treatment/control is ≥1.3-fold. Gene counts are based on unique gene symbols.

Treatments	No. DEGs	No. HPG-axis compiled genes in DEGs (gene symbols)	Treatments	No. DEGs	No. HPG-axis compiled genes in DEGS
1	105	2 (LDLR, MAPK8)	23	206	3 (CALM2B, CALM3A, STAR)
2	175	8 (ATP2A1L, EGR2A, LDLR, MAPK3, VTG1, VTG2, VTG3, VTG6)	25	114	1 (CALM2B)
3	162	2 (ZP2.3, ZP3B)	32	41	0
4	172	4 (AR, CYP4, HSD17B3, LDLR)	33	77	1 (CEBPD)
5	193	7 (ACVRL1, ATP2A2B, CYP17A1, ITGA2B, PDIA4, VTG1, VTG2)	34	98	1 (ESR2B)
6	865	21 (ATP2A2B, CASK, CEBPD, CTSD, CYP1A, CYP4, ESR1, HSD11B2, HSD17B12A, HSD17B3, IGF1RB, NS:ZF-A220, PLAA, PRKCBP1L, TH2, VTG1, VTG2, VTG3, VTG6, WU:FB64H05, ZP2.3)	35	132	0
7	196	3 (CACNA1C, ZP2.3, ZP3B)	36	196	2 (SP1, VLDLR)
8	105	1 (MAPK8)	37	136	1 (ESR2B)
9	157	3 (CACNA1C, CCNB1, HSD17B3)	38	159	0
10	224	3 (CACNA1C, FSHR, ZP2.3)	40	111	0
11	143	1 (CYP4)	41	124	0
13	42	0	43	200	4 (GHITM, IGF1, INHBAA, LDLR)
14	76	0	46	84	1 (CAMSAP1)
15	62	1 (ESR2B)	47	130	5 (ATP2B3, CYP2K6, HSD11B2, IGF1RA, NOS2)
16	134	2 (MAPKAPK5, PLA2G6)	49	52	1 (PLA2G6)
17	31	0	51	72	0
19	212	3 (CALM2B, CYP17A1, DHCR7)	52	75	0
20	142	3 (CACNA1C, HSD17B3, PDIP5)	54	206	2 (INHBAA, ZP3)
21	285	3 (CAMSAP1, CYP19A1A, IGF1RA)	56	347	9 (ATP2A1L, CACNA1C, CALM2B, CASK, CYP1A, HSD11B2, INHBAA, PDIA4, PLA2G6)
22	124	3 (CCNB1, CYP19A1A, STAR)	57	106	1 (FST)

All 10 test chemicals are known or hypothesized to interact with targets within the HPG-axis. But, only one treatment, EE2 48 h male testis, was found to have a significant effect on HPG-axis compiled as a group according to GSEA. When all 209 genes in HPG-axis compiled were mapped to the DEGs from 40 treatments where each set of DEG had a FDR < 40% and treatment/control ≥1.3-fold (Table 6), a total of 56 genes responded to 30 treatments involving all 10 chemicals. Measured by the number of HPG genes changed, the most influential treatments included EE2 48 h male liver, VIN 48 h male testis, and EE2 48 h male testis. Some of the HPG genes altered by the greatest number of treatments were CACNA1C, ZP2.3, LDLR, HSD17B3, CALM2B, VTG2, VTG1, PLA2G6, INHBAA, and HSD11B2. When 16 TF networks, representing 6 out of a total of 10 TFs from HPG-axis compiled, were individually evaluated, they were found significant for 15 treatments encompassing nine out of 10 chemicals (Table 7). The three most influential treatments, measured by number of networks per treatment, were TRI 96 h female ovary, EE2 48 h male liver, and EE2 48 h male testis.

Network clustering based on shared gene memberships could potentially shed additional light on biological mechanisms impacted by the putative EDCs. Of particular interest here are those TF networks and pathways overlapping, thus potentially interacting, with the HPG-axis, and those ranked high by various approaches, such as ATF4, CSDC2, TP53, p53 signaling, and TGF-beta signaling. In a dendrogram limited to HPG-axis compiled and KEGG pathways only, most of the chemical treatment-significant KEGG pathways by GSEA share substantial similarities to HPG-axis compiled. There are two distinct clusters juxtaposing HPG-axis compiled as neighbors. The first contains TGF-beta, p53 and a few other signaling pathways, and the other includes Jat–Stat, MAPK, and apoptosis (Fig. 2A). When all the TF networks were included with KEGG pathways and HPG-axis compiled in clustering, TP53 and XBP1 networks cluster with this same group of KEGG path-

Table 7
Sixteen TF networks representing six TFs in the HPG-axis compiled, and treatments under which they are significant at FDRs ≤ 0.25. Out of a total of 209 members in HPG-axis compiled, 10 are identified as TFs and mapped to 13 unique Agilent probes. If the networks of a hub TF built from multiple data sets are all significant, they are listed in an abbreviated format. For example, A_15_P100370 (ovaryS/ovaryR/testisS) represents DRTFovaryCy5Cy3_ A_15_P100370, DRTFovaryRatio_ A_15_P100370, and DRTFtestisCy5Cy3_ A_15_P100370, each of which is significant under its corresponding group of treatments delineated by a forward slash.

Significant TF networks	Annotations	Treatments
GSEA		
A_15_P100370 (ovaryS/ovaryR/testisS)	ESR1 estrogen receptor 1	6/2, 6/2, 8
A_15_P101177 (ovaryS/ovaryR/testisS)	CEBPG CCAAT/enhancer binding protein (C/EBP), gamma	17, 38, 50/17, 46, 50/17, 50
A_15_P109417 (ovaryS/testisS)	CEBPG	6, 33/6, 50
A_15_P108101 (testisS)	ESR2B estrogen receptor 2b	6
A_15_P109325 (ovaryS/testisS)	ESR2B	13/6, 21, 38, 50
A_15_P115095 (testisS)	ESR2B	6
A_15_P113550 (ovaryR/testisS)	AR androgen receptor	23, 38, 50/17, 38, 50
A_15_P119478 (testisS)	EGR2A early growth response 2a	2
A_15_P119871 (testisS)	EGR2B early growth response 2b	2, 6
E-GSEA		
A_15_P100370 (ovaryS/ovaryR/testisS)	ESR1	2/2/2
A_15_P101177 (ovaryS/testisS)	CEBPG	50/50
A_15_P109417 (ovaryS/testisS)	CEBPG	4, 6, 26/30
A_15_P109325 (testisS)	ESR2B	34
A_15_P113550 (testisS)	AR	50
A_15_P119871 (testisS)	EGR2B	2

ways (Supplemental Figure 1). The cluster containing HPG-axis compiled, however, is now composed of a number of TF networks headed primarily by TFs from gene families of HOX and FOX, and a few other nuclear receptors (Fig. 2B). Some of the top-ranked TF networks are also quite similar to one another (e.g., cluster with DRTFovaryCy5Cy3_A_15_P110418, Supplemental Figure 1), where closely related ATF4 and YY1 are clustered with such significant TF networks as TGIF1, SMAD5, CSDC2, and TFDP1. All matched genes between HPG-axis compiled and its closely related KEGG pathways/TF networks corresponding to Fig. 2A and B are provided in Supplemental Tables 1 and 2.

4. Discussion

The biological framework constructed in this study, made up of previously annotated signaling pathways and hundreds of reverse-engineered TF networks prioritized by their statistical associations with chemical treatments, represents a step forward toward studying gene regulatory mechanisms underlying chemical responses in zebrafish and related species. To overcome insufficient zebrafish genome annotations in terms of both gene identity and biological pathways, the framework was developed by employing three independent but complementary approaches, IPA mapping of DEGs to HMR pathways, GSEA/E-GSEA of networks and pathways, and UPGMA analysis of network/pathway similarities. Such a framework should contribute to the advancement of zebrafish in becoming a fully functional model system for toxicogenomics. Depending upon targets of interest, prioritization of pathways/networks could be conducted in a variety of ways to formulate new hypotheses for test chemicals, such as the HPG-active compounds assessed in this study.

The main goal of reverse-engineering TF networks in this study was to achieve reasonable accuracy in placing target genes into a given network so that the entire group of functionally related genes could be evaluated as a whole for responses to treatment conditions. Therefore, current issues in recovering correct gene topology/connectivities from constructing large networks (Hache et al., 2009; Kim et al., 2009) is not necessarily germane to the objectives of our study. Ideally still, the quality of TF networks should be assessed against a collection of empirically based, well-curated pathways. Two issues make such a quality assessment difficult for our study: almost half the Agilent zebrafish probes are not annotated and there are very few empirically constructed TF networks available to enable a robust evaluation. Consider, for example, SMAD and ESR, two TFs with well-understood pathways (TGF-beta signaling and estrogen receptor signaling). A total of 57 of 80 members of KEGG TGF-beta signaling pathway, some from multi-gene families, are mapped to Agilent probes. Since TGF-beta itself, or transforming growth factor, is not considered a TF, we turned to a group of SMADs which are both TFs and important components of the pathway. There should be a significant overlap between SMAD networks and TGF-beta signaling. Of the 22 SMAD networks constructed across the three data sets, the one overlapping the TGF-beta signaling pathway the most, DRTFovaryCy5Cy3_A_15_P100787, contains nine members of the pathway including BMPR1A, BMPR1B, CUL1A, RPS6KB1, SMAD1, SMAD5, TGFB3, TGFBR2, and BMP7B. Also found in this network are a few TFs known to be SMAD partners. In the case of ESR, only 9 out of 19 members of estrogen signaling pathway (as defined by GO:0030520) are mapped to Agilent probes. Among the 10 ESR1/ESR2 networks constructed, their only matches with the estrogen signaling pathway are the receptors themselves. While these two examples illustrate that TF networks reverse-engineered based upon microarray data likely vary in quality, the TF networks should still provide a method to formulate new hypotheses about potential MOAs of chemicals such as EDCs.

Several HMR pathways, ranked high for chemical responses as measured by number of treatments per significant pathway (Table 3), have extensive cross-talk. The IL-2, -3, -6, PI3K, and Jak–Stat pathways apparently all interact with one another (Rawlings et al., 2004). As a signaling pathway for many cytokines and growth factors, Jak–Stat activation promotes cell proliferation, differentiation, migration, and apoptosis. TGF-beta signaling shares similar functions as Jak–Stat (Dijke and Hill, 2004), and is regulated by Jak–Stat through interferon-gamma-activated STAT pathway on SMADs (Ulloa et al., 1999). Furthermore, PIAS, a key inhibitor of Jak–Stat, also regulates SMAD along with tumor suppressor p53 and androgen receptor (Shuai and Liu, 2003). Cellular responses to TGF-beta signaling rely on interactions between p53 family members and SMAD (Cordenonsi et al., 2003). Furthermore, in epithelial cells, TGF-beta induces apoptosis via p53 and a DNA-damage-sensitive kinase through SMAD7 (Zhang et al., 2006). P53 signaling coordinates cellular responses to a variety of stresses such as DNA damage, hypoxia, heat/cold shock, and oncogene activation, leading to cell cycle arrest, cellular senescence, and apoptosis (Hofseth et al., 2004; Levine et al., 2006). IGF-1 interfaces with p53 via MDM-2, an ubiquitin ligase. While signaling pathways involved in these fundamental cellular functions and developmental processes are conserved across animal species (Pires-daSilva and Sommer, 2003), there are still likely some signaling mechanisms unique to fish, and/or oviparous vertebrates in general, that distinguish the species biochemically, physiologically, and developmentally from mammals such as human, rat, and mouse (HMR). In this regard, zebrafish-specific TF networks constructed in this study and associated with EDC chemicals could provide additional insights into toxicity mechanisms that are more specific and germane to fish species than the HMR pathways.

The top findings from GSEA and E-GSEA, as ranked by number of treatments per significant network, are enriched with many TFs known to be mutually interactive. They are primarily involved in cellular functions such as cellular stress, cell cycle control, and apoptosis, and thus largely consistent with the significant HMR pathways described above. The KEGG p53 signaling pathway was quite responsive to the test chemicals, as were the networks of TP53 (an alias of p53), one of the two identified TFs in the p53 pathway. A total of 12 different treatments from 7 chemicals have a significant impact on TP53 networks, with EE2 and FLU the most common. Both ATF4 and XBP1 are members of CREB/ATF family (cAMP response element binding/activating transcription factor), a class of TFs containing the basic leucine zipper (b-zip) structure. ATF4 is a member of both GnRH and MAPK signaling pathways (http://www.genome.jp/kegg/) and an interacting partner to many proteins including the $GABA_B$ receptor (Ritter et al., 2004). ATF4 is induced by stress signals such as anoxia/hypoxia, endoplasmic reticulum (ER) stress, and oxidative stress (Ameri and Harris, 2008) and involved in fertility and cell proliferation and differentiation. Like ATF4, XBP1 is also involved in ER stress by activating the unfolded protein response pathway (Koong et al., 2006). In humans and yeast, XBP1 interacts with both ATF4 and ATF6 through CREBZF, another b-zip class TF. Additional high-ranked TFs sharing similar functions in cell cycle and apoptosis include DPF2 (Gabig et al., 1994), MYBL2 (Sala, 2005), MEF2D (Potthoff and Olson, 2007), and TFDP1 (Hitchens and Robbins, 2003). TFDP1 is a member of both cell cycle and TGF-beta signaling in KEGG annotations. TFDP1-mediated apoptosis may be p53-dependent or -independent. Some of these hub TFs from the top-ranked TF networks interact with one another. TGIF1 is a co-repressor of SMAD (Wotton et al., 1999). YY1 is a ubiquitous and multifunctional TF fundamental to embryogenesis, cell differentiation and proliferation (Gordon et al., 2006). It shares sequence homology to KLF2B and KLF15, *Drosophila* Krup-

pel proteins with C2H2 zinc finger motifs. YY1 interacts with a variety of proteins, including YBX1, ATF, p53, SP1, SMAD, and TGF-beta. CSDC2 and YBX1 are both members of cold shock domain protein family, which bind to both DNA and RNA (Matsumoto and Wolffe, 1998). While CSDC2 is relatively poorly characterized, YBX1 is expressed throughout embryogenesis and ubiquitously in adult animals. It interacts with SMAD3 and p53 among other genes, and plays a role in oncogenesis (Wu et al., 2007). Coincidentally, YBX1 is also one of the most networked genes among the TF networks constructed in this study. And activity/stability of YY1, p53, androgen receptor, estrogen receptor, SMAD7, STAT3 (Jak–Stat signaling), and E2F (a partner of TFDP1) could all be modulated through protein acetylation/deacetylation (Glozak et al., 2005). Clearly, similar to HMR pathways discussed earlier, TF networks identified by GSEA/E-GSEA and ranked high in this prioritization scheme also tend to be mutually interactive and share similar cellular functions, thus offering insights into gene regulatory mechanisms underlying biological responses to exposure of these chemicals.

The degree of impact of the 10 putative EDCs used in this study on known HPG-axis compiled genes (Villeneuve et al., 2007) varied based on the analysis considered. According to GSEA, the chemicals seemed to have little impact on HPG-axis compiled as a whole. The FDR rates for HPG-axis compiled in GSEA exceeded 0.5 in 54 out of the 58 treatments. However, we note that the full complement of 209 genes associated with HPG-axis compiled represents various pathways, processes, and regulatory features that are spatially separated from one another in four distinct tissue compartments, as well as distinct cell types within those compartments. Therefore, we would not really expect to see significant impacts on this noncohesive group of genes as a whole within the narrow scope of chemical/dose/duration/tissue captured by each individual treatment. However, an examination of individual genes from HPG-axis compiled resulted in quite a different picture. While on average very few members of HPG-axis compiled responded to any given chemical treatment, likely as a result of low statistical power due to a small sample size (Pawitan et al., 2005), in combination a total of 56 HPG genes, or 27% of the group, were significantly changed by 30 treatments involving all 10 test chemicals. When the networks of all 10 hub TFs contained in HPG-axis compiled were examined individually by GSEA and E-GSEA, 6 of them were changed by 15 treatments involving all the chemicals except VIN. Interestingly, the 56 differentially expressed HPG genes contained only 4 of the 10 TFs. None of the top 10 DEGs among the 56, as ranked by number of treatments per DEG, was a TF and each of these 10 DEGs responded to 3–5 treatments. Since the HPG-axis really should be viewed as a network made up of many peripheral sub-networks continuously branching-out and operating in multiple different tissues, evaluations of chemical impacts by GSEA are obviously highly dependent on what criteria are used to determine the gene memberships of HPG-axis compiled. Overall, our findings suggest that HPG-axis compiled is substantially impacted by the 10 EDCs tested in this study, and that the members of HPG-axis compiled are not equally affected by the chemicals. Further, our results indicate that some of its peripheral TF networks may be specifically targeted by the test chemicals.

In a three-dimensional network space, networks with shared gene nodes should have some functional interactions, or cross-talk. When measured by Jaccard distance based on the degree of node-sharing, those networks with a greater overlap should become closer neighbors in a dendrogram. There are many instances of this relationship present in the dendrograms constructed in this study for TF networks and KEGG pathways. Given the impact of the 10 test chemicals on HPG-axis compiled, p53 and TGF-beta signaling pathways, perhaps it should not be surprising to find that, when limited to only HPG-axis compiled and KEGG pathways, the two clusters closest to HPG-axis compiled are represented by TGF-beta/p53 and a few other pathways known to be involved in cell cycle, stress responses, and apoptosis. In fact, many of these neighboring KEGG pathways have extensive cross-talk among themselves in order to exert their regulation of, and impact on, these cellular functions (Guo and Wang, 2009). The androgen receptor (AR) and estrogen receptor (ESR1), both TFs in HPG-axis compiled, also interact with TGF-beta through SMAD proteins (Moustakas et al., 2001). Interestingly, when considered in the context of an all-inclusive dendrogram of 1932 nodes (Fig. 2B, Supplemental Figure 1), the TF networks closest to HPG-axis compiled are mostly headed by TFs with known or potential involvement in endocrine responses. Both FIGLA and POU5F1 are oocyte-specific genes critical to early folliculogenesis or embryogenesis (Pangas and Rajkovic, 2006). While FOXH1 interacts with SMAD to mediate TGF-beta signals (Feng and Derynck, 2005), FOXA proteins appear to be partners of nuclear hormone receptors (Hannenhalli and Kaestner, 2009). And, the expression of HOX genes is regulated by several hormones and their receptors, and disrupted by EDCs (Daftary and Taylor, 2006). One final example of cross-talk features some of the TF networks ranked high in our analysis, ATF4, YY1, TGIF1, SMAD5, PIAS1, CSDC2, and TFDP1. As discussed above, the interactions among many of these TFs have been documented in literature. Again, they are located in the same cluster. All of these examples strongly suggest that clustering based on the similarities among TF networks/pathways could serve as a powerful tool to explore biological mechanisms underlying responses to endocrine-active chemicals.

5. Conclusions

Using several independent and complementary approaches, we have constructed a transcriptomics-based biological framework made up of hundreds of statistically prioritized signaling pathways and TF networks for studying biological responses to chemicals known to affect endocrine function. While the quality of these TF networks may vary due to limitations in the algorithms of reverse-engineering and sole dependence on gene transcriptional profiling data, we have demonstrated the utility of this framework in formulating new hypotheses about EDC toxicity mechanisms. The potential biological pathways underlying the tested chemicals have been expanded from initially a single, somewhat subjectively defined HPG-axis compiled to several well-curated KEGG pathways including TGF-beta, p53 signaling, and their cross-talking network neighbors, and to a still greater number of tentative TF networks peripheral to them, all of which could have a profound impact on cellular functions such as stress response, cell cycle, and apoptosis. The increased specificity among these testable hypothetical toxicity pathways could be utilized in the future to design experiments focusing on various aspects of EDC mechanisms. Furthermore, these putative pathways and their hub TFs could potentially provide candidates for the much-needed mechanism-based biomarkers. The mutants of genes in these pathways would facilitate the discovery and development of more specific, mechanism-based end points relevant to risk assessment. Perhaps just as significant, the biological framework constructed in this study would also greatly reduce gene dimensionality, thus bypassing the dilemma of small sample size/large number of variables encountered in typical toxicogenomics studies. And, it could be further improved in the future by adding more samples perturbed under a greater variety of conditions, using more sophisticated reverse-engineering algorithms, and having better genome annotations.

Acknowledgements

The US Environmental Protection Agency (US EPA) through its Office of Research and Development funded and managed the

research described here. The paper has been subjected to Agency's administrative review and approved for publication as a U.S. EPA document. This work was partially supported by an award from the National Center for Computational Toxicology to the Ecological Exposure Research and Ecosystem Research Divisions (National Exposure Research Laboratory) in Cincinnati, Ohio, and Athens, Georgia, USA, respectively, and the Mid-Continent Ecology Division (National Health and Environmental Effects Research Laboratory) in Duluth, MN, USA. We wish to thank Wei Keat Lim and Manjunath Kustagi from the Andrea Califano lab at the Center for Computational Biology and Bioinformatics, Columbia University, New York city, USA for advice in implementing E-GSEA method and conducting ARACNE analysis. Timely computing support was provided by Edward Anderson and Robert McCauley of the US EPA National Computer Center at Research Triangle Park, North Carolina. The manuscript has greatly benefited from critiques by two anonymous reviewers, and Lyle Burgoon at the US EPA National Health and Environmental Effects Research Laboratory at Research Triangle Park, North Carolina.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.aquatox.2010.02.021.

References

Ameri, K., Harris, A.L., 2008. Activating transcription factor 4. Int. J. Biochem. Cell. Biol. 40, 14–21.

Ankley, G.T., Villeneuve, D.L., 2006. The fathead minnow in aquatic toxicology: past, present and future. Aquat. Toxicol. 78, 91–102.

Ankley, G.T., Bencic, D.C., Breen, M.S., Collette, T.W., Conolly, R.B., Denslow, N.D., Edwards, S.W., Ekman, D.R., Garcia-Reyero, N., Jensen, K.M., Lazorchak, J.M., Martinovic, D., Miller, D.H., Perkins, E.J., Orlando, E.F., Villeneuve, D.L., Wang, R.-L., Watanabe, K.H., 2009. Endocrine disrupting chemicals in fish: developing exposure indicators and predictive models of effects based on mechanism of action. Aquat. Toxicol. 92, 168–178.

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ. Toxicol. Chem. 29, 730–741.

Bansal, M., Belcastro, V., Ambesi-Impiombato, A., di Bernardo, D., 2007. How to infer gene networks from expression profiles. Mol. Syst. Biol. 3, 78.

Baralla, A., Mentzen, W.I., de la Fuente, A., 2009. Inferring gene networks: dream or nightmare? Part 1: challenges 1 and 3. Ann. N. Y. Acad. Sci. 1158, 246–256.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera1, R., Califano, A., 2005. Reverse engineering of regulatory networks in human B cells. Nat. Genet. 37, 382–390.

Benjamini, B., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300.

Blais, A., Dynlacht, B.D., 2005. Constructing transcriptional regulatory networks. Genes Dev. 19, 1499–1511.

Chen, K., Rajewsky, N., 2007. The evolution of gene regulation by transcription factors and microRNAs. Nat. Rev. Genet. 8, 93–103.

Choi, S.M., Lee, B.M., 2004. An alternative mode of action of endocrine-disrupting chemicals and chemoprevention. J. Toxicol. Environ. Health 7, 451–463.

Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., Piccolo, S., 2003. Links between tumor suppressors: p53 is required for tgf-beta gene responses by cooperating with Smads. Cell 113, 301–314.

Daftary, G., Taylor, H., 2006. Endocrine regulation of HOX genes. Endocr. Rev. 27, 331–355.

Dijke, P., Hill, C.S., 2004. New insights into TGF-b–Smad signalling. Trends Biochem. Sci. 29, 265–273.

ECETOC, 2007. Intelligent testing strategies in ecotoxicology: mode of action approach for specifically acting chemicals. In: ECETOC Technical Report No. 102. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium, http://www.ecetoc.org/technical-reports.

Edwards, S., Preston, R., 2008. Systems biology and mode of action based risk assessment. Toxicol. Sci. 106, 312–318.

Feng, X.-H., Derynck, R., 2005. Specificity and versatility in tgf-β signaling through Smads. Annu. Rev. Cell Dev. Biol. 21, 659–693.

Gabig, T.G., Mantel, P.L., Rosli, R., Crean, C.D., 1994. Requiem: a novel zinc finger gene essential for apoptosis in myeloid cells. J. Biol. Chem. 269, 29515–29519.

Glozak, M.A., Sengupta, N., Zhang, X.-H., Seto, E., 2005. Acetylation and deacetylation of non-histone proteins. Gene 363, 15–23.

Gordon, S., Akopyan, G., Garban, H., Bonavida, B., 2006. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. Oncogene 25, 1125–1142.

Guo, X., Wang, X.-F., 2009. Signaling cross-talk between TGF-β/BMP and other pathways. Cell Res. 19, 71–88.

Hache, H., Lehrach, H., Herwig, R., 2009. Reverse engineering of gene regulatory networks: a comparative study. EURASIP J. Bioinform. Syst. Biol., doi:10.1155/2009/617281.

Hannenhalli, S., Kaestner, K., 2009. The evolution of Fox genes and their role in development and disease. Nat. Rev. Genet. 10, 233–240.

Hauser, R., Barthold, J., Meeker, J., 2007. Epidemiologic evidence on the relationship between environmental endocrine disruptors and male reproductive and developmental health. In: Gore, A. (Ed.), Endocrine Disrupting Chemicals. Humana Press, New Jersey, pp. 225–251.

Hitchens, M.R., Robbins, P.D., 2003. The role of the transcription factor DP in apoptosis. Apoptosis 8, 461–468.

Hofseth, L.J., Hussain, S.P., Harris, C.C., 2004. p53: 25 years after its discovery. Trends Pharmacol. Sci. 25, 177–181.

Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., Rupp, R., 2007. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics 8, 460–466.

Jaccard, P., 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Société Vaudoise des Sciences Naturelles 37, 547–579.

Janssen, S., Fujimoto, V., Giudice, L., 2007. Endocrine disruption and reproductive outcomes in women. In: Gore, A. (Ed.), Endocrine Disrupting Chemicals. Humana Press, New Jersey, pp. 203–223.

Kavlock, R.J., Daston, G.P., DeRosa, C., Fenner-Crisp, P., Gray, L.E., Kaattari, S., Lucier, G., Luster, M., Mac, M.J., Maczka, C., Miller, R., Moore, J., Rolland, R., Scott, G., Sheehan, D.M., Sinks, T., Tilson, H.A., 1996. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: a report of the U.S. EPA-sponsored workshop. Environ. Health Perspect. 104, 715–740.

Kidd, K.A., Blanchfield, P.J., Mills, K.H., Palace, V.P., Evans, R.E., Lazorchak, J.M., Flick, R.W., 2007. Collapse of a fish population after exposure to a synthetic estrogen. Proc. Natl. Acad. Sci. U. S. A. 104, 8897–8901.

Kim, H., Shay, D., O'Shea, T., Regev, E.K.A., 2009. Transcriptional regulatory circuits: predicting numbers from alphabets. Science 325, 429–432.

Koong, A.C., Chauhan, V., Romero-Ramirez, L., 2006. Targeting XBP-1 as a novel anticancer strategy. Cancer Biol. Ther. 5, 756–759.

Levine, A.J., Hu, W., Feng, Z., 2006. The P53 pathway: what questions remain to be explored? Cell Death Differ. 13, 1027–1036.

Levine, M., Tjian, R., 2003. Transcription regulation and animal diversity. Nature 424, 147–151.

Lim, W.K., Eugenia, L., Califano, A., 2009. Master regulators used as breast cancer metastasis classifier. Pac. Symp. Biocomput. 14, 504–515.

Manning, T., 2005. Endocrine-disrupting chemicals: a review of the state of the science. Aust. J. Ecotoxicol. 11, 1–52.

Margolin, A., Wang, K., Lim, W., Kustagi, M., Nemenman, I., Califano, A., 2006. Reverse engineering cellular networks. Nat. Protoc. 1, 663–672.

Matsumoto, K., Wolffe, A.P., 1998. Gene regulation by Y-box proteins:coupling control of transcription and translation. Trends Cell Biol. 8, 318–323.

Melnick, R., Lucier, G., Wolfe, M., Hall, R., Stancel, G., Prins, G., Gallo, M., Reuhl, K., Ho, S.M., Brown, T., Moore, J., Leakey, J., Haseman, J., Kohn, M., 2002. Summary of the national toxicology program's report of the endocrine disruptors low-dose peer review. Environ. Health Perspect. 110, 427–431.

Moustakas, A., Souchelnytskyi, S., Heldin, C.H., 2001. Smad regulation in TGF-b signal transduction. J. Cell Sci. 114, 4359–4369.

National Research Council (NRC), 2007. Toxicity Testing in the Twenty-first Century: A Vision and A Strategy. National Academy of Sciences, Washington, DC, USA.

Pangas, S.A., Rajkovic, A., 2006. Transcriptional regulation of early oogenesis: in search of masters. Hum. Reprod. Update 12, 65–76.

Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A., Ploner, A., 2005. False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics 21, 3017–3024.

Pires-daSilva, A., Sommer, R., 2003. The evolution of signaling pathways in animal development. Nat. Rev. Genet. 4, 39–49.

Potthoff, M.J., Olson, E.N., 2007. MEF2: a central regulator of diverse developmental programs. Development 134, 4131–4140.

Rawlings, J.S., Rosler, K.M., Harrison, D.A., 2004. The JAK/STAT signaling pathway. J. Cell Sci. 117, 1281–1283.

Ritter, B., Zschuntsch, J., Kvachnina, E., Zhang, W., Ponimaskin, E.G., 2004. The GABA(B) receptor subunits R1 and R2 interact differentially with the activation transcription factor ATF4 in mouse brain during the postnatal development. Brain Res. Dev. Brain Res. 149, 73–77.

Rottmann, S., Lüscher, B., 2006. The Mad side of the Max network: antagonizing the function of Myc and more. Curr. Top. Microbiol. Immunol. 302, 63–122.

Sala, A., 2005. B-MYB, a transcription factor implicated in regulating cell cycle, apoptosis and cancer. Eur. J. Cancer 41, 2479–2484.

Sanderson, J., 2006. The steroid hormone biosynthesis pathway as a target for endocrine-disrupting chemicals. Toxicol. Sci. 94, 3–21.

Scheinine, A., Mentzen, W.I., Fotia, G., Pieroni, E., Maggio, F., Mancosu, G., de la Fuente, A., 2009. Inferring gene networks: dream or nightmare? Part 2: challenges 4 and 5. Ann. N. Y. Acad. Sci. 1158, 287–301.

Shuai, K., Liu, B., 2003. Regulation of Jak–Stat signalling in the immune system. Nat. Rev. Immunol. 3, 900–911.

Sprague, J., Bayraktaroglu, L., Clements, D., Conlin, T., Fashena, D., Frazer, K., Haendel, M., Howe, D., Mani, P., Ramachandran, S., Schaper, K., Segerdell, E., Song, P., Sprunger, B., Taylor, S., Van Slyke, C., Westerfield, M., 2006. The Zebrafish Information Network: the zebrafish model organism database. Nucleic Acids Res. 34, D581–585.

Subramanian, A., Tamayoa, P., Moothaa, V.K., Mukherjeed, S., Eberta, B.L., Gillettea, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene Set Enrichment Analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550.

Tabb, M., Blumberg, B., 2006. New modes of action for endocrine-disrupting chemicals. Mol. Endocrinol. 20, 475–482.

Takao, M., Oohata, Y., Kitadokoro, K., Kobayashi, K., Iwai, S., Yasui, A., Yonei, S., Zhang, Q.M., 2009. Human Nei-like protein NEIL3 has AP lyase activity specific for single-stranded DNA and confers oxidative stress resistance in *Escherichia coli* mutant. Genes Cells 14, 261–270.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B., 2001. Missing value estimation methods for DNA microarrays. Bioinformatics 17, 520–525.

Ulloa, L., Doody, J., Massaguae, J., 1999. Inhibition of transforming growth factor-β/SMAD signalling by the interferon-γ/STAT pathway. Nature 397, 710–713.

US Environmental Protection Agency, 1998. Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) Final Report. Office of Prevention, Pesticides and Toxic Substances, US Environmental Protection Agency, Washington, DC, http://www.epa.gov/scipoly/oscpendo/pubs/edspoverview/finalrpt.htm.

Vaquerizas, J., Kummerfeld, S., Teichmann, S., Luscombe, N., 2009. A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 10, 252–263.

Villeneuve, D.L., Larkin, P., Knoebl, I., Miracle, A.L., Kahl, M.D., Jensen, K.M., Makynen, E.A., Durhan, E.J., Carter, B.J., Denslow, N.D., Ankley, G.T., 2007. A graphical systems model to facilitate hypothesis-driven ecotoxicogenomics research on the brain–pituitary–gonadal axis. Environ. Sci. Technol. 40, 321–330.

Villeneuve, D.L., Wang, R.-L., Bencic, D.C., Biales, A.D., Martinovic, D., Lazorchak, J.M., Toth, G., Ankley, G.T., 2009. Altered gene expression in the brain and ovaries of zebrafish exposed to the aromatase inhibitor fadrozole: microarray analysis and hypothesis generation. Environ. Toxicol. Chem. 28, 1767–1782.

Wang, R.-L., Biales, A.D., Bencic, D.C., Lattier, D., Kostich, M., Villeneuve, D.L., Ankley, G.T., Lazorchak, J.M., Toth, G., 2008a. DNA microarray application in ecotoxicology: experimental design, microarray scanning, and factors affecting transcriptional profiles in a small fish species. Environ. Toxicol. Chem. 27, 652–663.

Wang, R.-L., Bencic, D.C., Biales, A.D., Lattier, D., Kostich, M., Villeneuve, D.L., Ankley, G.T., Lazorchak, J.M., Toth, G., 2008b. DNA microarray-based ecotoxicological biomarker discovery in a small fish model species. Environ. Toxicol. Chem. 27, 664–675.

Werhli, A.V., Grzegorczyk, M., Husmeier, D., 2006. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks. Bioinformatics 22, 2523–2531.

Wilson, D., Charoensawan, V., Kummerfeld, S.K., Teichmann, S.A., 2008. DBD—taxonomically broad transcription factor predictions: new content and functionality. Nucleic Acids Res. 36, D88–92.

World Health Organization (WHO), 2002. Global assessment of the state-of-the-science of endocrine disruptors. In: Damstra, T., Barlow, S., Bergman, A., Kavlock, R., Van Der Kraak, G. (Eds.), International Programme on Chemical Safety, WHO/PCS/EDC/02.2.

Wotton, D., Lo, R.S., Lee, S., Massague, J., 1999. A Smad transcriptional corepressor. Cell 97, 29–39.

Wu, J., Stratford, A.L., Astanehe, A., Dunn, S.E., 2007. YB-1 is a transcription/translation factor that orchestrates the oncogenome by hardwiring signal transduction to gene expression. Translat. Oncogenomics 2, 49–65.

Zeeberg, B.R., Qin, H.-Y., Narasimhan, S., Sunshine, M., Cao, H., Kane, D.W., Reimers, M., Stephens, R.M., Bryant, D., Burt, S.K., Elnekave, E., Hari, D.M., Wynn, T.A., Cunningham-Rundles, C., Stewart, D.M., Nelson, D., Weinstein, J.N., 2005. High-throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of common variable immune deficiency (CVID). BMC Bioinformatics 6, 168–186.

Zhang, S., Ekman, M., Thakur, N., Bu, S., Davoodpour, P., Grimsby, S., Tagami, S., Heldin, C.H., Landström, M., 2006. TGF beta1-induced activation of ATM and p53 mediates apoptosis in a Smad7-dependent manner. Cell Cycle 5, 2787–2795.